



A recursive approach to detect multivariable conditional variance components and conditional random effects

Jixiang Wu^a, Dongfeng Wu^b, Johnnie N. Jenkins^{c,*},
Jack C. McCarty Jr.^c

^aDepartment of Plant and Soil Sciences, Mississippi State University, Mississippi State, MS 39762, USA

^bDepartment of Mathematics and Statistics, Mississippi State University, Mississippi State, MS 39762, USA

^cCrop Science Research Laboratory, USDA-ARS, P. O. Box 5367, Mississippi State, MS 39762, USA

Received 1 December 2003; received in revised form 1 August 2004; accepted 12 August 2004

Available online 11 September 2004

Abstract

A complex trait like crop yield is determined by its component traits. Multivariable conditional analysis in a general mixed linear model is helpful in dissecting the gene expression for the complex trait due to different effects, such as environment, genotype, and genotype \times environment interaction. A recursive approach is presented for constructing a new random vector that can be equivalently used to analyze multivariable conditional variance components and conditional effects. End-of-season plant mapping data, including lint yield and three yield components for nine cultivars of upland cotton (*Gossypium hirsutum* L.) were used to detect the conditional variance components and conditional effects using this new approach, which can help identify genotypes to be used in selection studies.

© 2004 Elsevier B.V. All rights reserved.

1. Introduction

Gene expression for a complex trait depends on the joint contribution of its component traits, as well as other factors. In crops, for example, a complex trait like yield is determined by several yield components. The existence of correlations between a complex trait and its

* Corresponding author. Tel.: +1-6623207386; fax: +1-6623207528.

E-mail address: jjenkins@msa-msstate.ars.usda.gov (J.N. Jenkins).

components could be an indication of gene association or pleiotropy (Kebede et al., 2001; Dilday et al., 1990).

Correlation analysis, multiple linear regression analysis, and path analysis have been commonly used to detect the relationships between a complex trait and its components (Bora et al., 1998; Ball et al., 2001; Cramer and Wehner, 2000; Samonte et al., 1998). Correlation analysis can only detect simple phenotypic relationship between a complex trait and each of its component traits, rather than the relative contribution of each individual component trait. Multiple linear regression analysis can reveal single or joint contributions of component traits to the complex trait. Path analysis (Wright, 1920) can divide the simple correlation coefficients into direct and indirect effects of component traits on the target trait. Both path analysis and multiple linear regression analysis have several shortcomings. First, the magnitudes of contribution coefficients are influenced by the number of component traits for both methods. Second, multiple linear regression analysis is based on the assumption that all explanatory variables are fixed, but, in most cases, both the complex trait and its component traits are random. Thus, regarding the component traits as explanatory variables, ignoring their random error in many multiple linear regression analyses, violates this basic assumption. Jobson (1991) proposed the conditional multiple linear regression method based on multiple normal condition distribution theory. In an experimental design or a complex genetic design, both a complex trait and its component traits may follow the same mixed linear model; however, these methods cannot partition phenotypic contribution from its component traits into different parts due to genetic, or environmental effects. Using a mixed linear model approach, the total phenotypic covariance can be partitioned into different covariance components between pairs of traits (Zhu, 1989; Wu et al., 1995). This, however, can only detect relationships due to effects such as genetic and non-genetic parts.

Conditional probability models can be used to detect extra variation (conditional variances) without the influence of the given variables (Graybill, 1976; Krzanowski, 1988). Based on the conditional variance, the variation due to the given variables can be obtained. The expectation and the variance–covariance matrix for a conditional random vector can be derived (Graybill, 1976; Krzanowski, 1988); however, the conditional variance components in a mixed linear model are not directly obtainable. Zhu (1995) proposed a method that could be used to analyze the conditional effects and the conditional variance components to explore gene expression for single developmental traits. This conditional model approach has been applied in the study of developmental genetics in crops (Zhu, 1995; Yan et al., 1998) and mice (Atchley and Zhu, 1997). The conditional model approach also helps to analyze the contribution of each component trait to a complex trait (Zhu, 1995); however, this conditional approach can only analyze single-variable conditional variance components and conditional random effects. According to multivariate distribution theory, multivariable conditional expectation and variance–covariance matrix can be easily derived by combining mixed linear approaches and classical methods (Graybill, 1976; Krzanowski, 1988); however, the multivariable conditional random effects and variance components still cannot be derived directly. It is also computationally difficult to directly use the traditional method (Graybill, 1976; Krzanowski, 1988) and the conditional model approach (Zhu, 1995) to analyze multivariable conditional variance components and random effects, especially for complicated genetic models with large sample size. An approach that can analyze multi-

variable conditional variance components and random effects under a general mixed linear model with less computational intensity would be highly desirable.

A recursive approach is presented for deriving a random vector that can be equivalently used to detect multivariable conditional variance components and random effects under a general mixed linear model. Statistical methods are also suggested for multivariable conditional analysis. Data for lint yield and three lint yield components of upland cotton are used to illustrate this new method.

2. Methodology

2.1. Derivation of independent variables

Suppose there are two normally distributed vectors that follow the same mixed linear model:

$$\mathbf{y}_{(i)} = \mathbf{X}\mathbf{b}_{(i)} + \sum_{u=1}^r \mathbf{U}_u \mathbf{e}_{u(i)} \sim N \left(\mathbf{X}\mathbf{b}_{(i)}, \mathbf{V}_{(i)} = \sum_{u=1}^r \sigma_{u(i)}^2 \mathbf{U}_u \mathbf{U}_u^T \right), \quad i = 1, 2, \quad (1)$$

where, $\mathbf{y}_{(i)}$ is an $n \times 1$ observations on response variable i , \mathbf{X} is the design information matrix for the fixed effect vector $\mathbf{b}_{(i)}$, \mathbf{U}_u is the design matrix for the random effect vector $\mathbf{e}_{u(i)}$, $\mathbf{e}_{u(i)} \sim N(0, \sigma_{u(i)}^2 \mathbf{I}_u)$.

Suppose $\begin{pmatrix} \mathbf{y}_{(1)} \\ \mathbf{y}_{(2)} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{X}\mathbf{b}_{(1)} \\ \mathbf{X}\mathbf{b}_{(2)} \end{pmatrix}, \Sigma_{(1,2)} \right)$, where $\Sigma_{(1,2)} = \begin{pmatrix} \mathbf{V}_{(1)} & \mathbf{C}_{(1,2)} \\ \mathbf{C}_{(2,1)} & \mathbf{V}_{(2)} \end{pmatrix}$, and $\mathbf{C}_{(1,2)}$ is the covariance matrix between $\mathbf{y}_{(1)}$ and $\mathbf{y}_{(2)}$ and is expressed as follows:

$$\mathbf{C}_{(1,2)} = \sum_{u=1}^r \sigma_{u(1,2)} \mathbf{U}_u \mathbf{U}_u^T \quad (2)$$

where $\sigma_{u(1,2)}$ is the covariance component between $\mathbf{y}_{(1)}$ and $\mathbf{y}_{(2)}$ for u th random effect. The covariance components between pairs of traits can be estimated by the method suggested by [Zhu and Weir \(1994\)](#).

The conditional distribution of $\mathbf{y}_{(2)}$, given $\mathbf{y}_{(1)}$, is defined as $\mathbf{y}_{(2|1)}$, which can be expressed in terms of matrices and vectors as

$$\mathbf{y}_{(2|1)} = \mathbf{X}\mathbf{b}_{(2|1)} + \sum_{u=1}^r \mathbf{U}_u \mathbf{e}_{u(2|1)}, \quad (3)$$

$\mathbf{y}_{(2|1)}$ is also normally distributed with mean

$$\mathbf{X}\mathbf{b}_{(2)} + \mathbf{C}_{(2,1)} \mathbf{V}_{(1)}^{-1} (\mathbf{y}_{(1)} - \mathbf{X}\mathbf{b}_{(1)}) \quad (4)$$

and variance–covariance matrix

$$\mathbf{V}_{(2)} - \mathbf{C}_{(2,1)} \mathbf{V}_{(1)}^{-1} \mathbf{C}_{(1,2)} = \sum_{u=1}^r \sigma_{u(2|1)}^2 \mathbf{U}_u \mathbf{U}_u^T, \quad (5)$$

where, $\mathbf{e}_{u(2|1)}$ and $\sigma_{u(2|1)}^2$ are the conditional vector random effects and variance component for the u th random effect, respectively. Based on Eq. (5), the conditional variance and

covariance matrix of $\mathbf{y}_{(2)}$, given $\mathbf{y}_{(1)}$, is clear; however, the conditional variance components and the conditional random effects cannot be derived. A new vector was suggested by [Zhu \(1995\)](#) for analyzing conditional effects and their conditional variance components.

$$\text{Let } \mathbf{y}_{(2|1)}^* = \mathbf{y}_{(2)} - \mathbf{C}_{(2,1)} \mathbf{V}_{(1)}^{-1} (\mathbf{y}_{(1)} - \mathbf{X}\mathbf{b}_{(1)}), \quad (6)$$

which is independent of \mathbf{y}_1 with mean of

$$\mathbf{X}\mathbf{b}_{(2|1)}^* = E(\mathbf{y}_{(2|1)}^*) = \mathbf{X}\mathbf{b}_{(2)} - \mathbf{C}_{(2,1)} \mathbf{V}_{(1)}^{-1} (\mathbf{y}_{(1)} - \mathbf{X}\mathbf{b}_{(1)}) \quad (7)$$

and the same variance–covariance matrix as that of $\mathbf{y}_{(2)}$, given $\mathbf{y}_{(1)}$ (see appendix).

The new vector $\mathbf{y}_{(2|1)}^*$ can be expressed by the same mixed linear model in Eq. (1), as follows:

$$\mathbf{y}_{(2|1)}^* = \mathbf{X}\mathbf{b}_{(2|1)}^* + \sum_{u=1}^r \mathbf{U}_u \mathbf{e}_{u(2|1)}^* \sim N \left(\mathbf{X}\mathbf{b}_{(2|1)}^*, \mathbf{V}_{(2|1)}^* = \sum_{u=1}^r \sigma_{u(2|1)}^{*2} \mathbf{U}_u \mathbf{U}_u^T \right), \quad (8)$$

$\sigma_{u(2|1)}^{*2}$ and $\mathbf{e}_{u(2|1)}^*$ in (8) are equivalent to the conditional variance components $\sigma_{u(2|1)}^2$ and the conditional effects $\mathbf{e}_{u(2|1)}$, respectively.

Now we extend the result to the case of p ($p > 2$) random vectors. If the p ($p > 2$) random vectors have a multiple joint normal distribution and each follows the same mixed linear model, the conditional vector $\mathbf{y}_{(p)}$ on $\mathbf{y}_{(1)}, \mathbf{y}_{(2)}, \dots, \mathbf{y}_{(p-1)}$ is defined as $\mathbf{y}_{(p|1,2,\dots,p-1)}$, which can be expressed in terms of matrices and vectors,

$$\mathbf{y}_{(p|1,2,\dots,p-1)} = \mathbf{X}\mathbf{b}_{(p|1,2,\dots,p-1)} + \sum_{u=1}^r \mathbf{U}_u \mathbf{e}_{u(p|1,2,\dots,p-1)}, \quad (9)$$

where $\mathbf{y}_{(p|1,2,\dots,p-1)}$ is normally distributed with expectation

$$\begin{aligned} E(\mathbf{y}_{(p|1,2,\dots,p-1)}) &= \mathbf{X}\mathbf{b}_{(p)} + (\mathbf{C}_{(p,1)} \mathbf{C}_{(p,2)} \dots \mathbf{C}_{(p,p-1)}) \\ &\times \sum_{(1,2,\dots,p-1)}^{-1} \begin{pmatrix} \mathbf{y}_{(1)} - \mathbf{X}\mathbf{b}_{(1)} \\ \dots \\ \mathbf{y}_{(p-1)} - \mathbf{X}\mathbf{b}_{(p-1)} \end{pmatrix} \end{aligned} \quad (10)$$

and variance–covariance matrix

$$\begin{aligned} \mathbf{V}_{(p|1,2,\dots,p-1)} &= \mathbf{V}_{(p)} - (\mathbf{C}_{(p,1)} \mathbf{C}_{(p,2)} \dots \mathbf{C}_{(p,p-1)}) \\ &\times \sum_{(1,2,\dots,p-1)}^{-1} (\mathbf{C}_{(1,p)} \mathbf{C}_{(2,p)} \dots \mathbf{C}_{(p-1,p)})^T \\ &= \sum_{u=1}^r \sigma_{u(p|1,2,\dots,p-1)}^2 \mathbf{U}_u \mathbf{U}_u^T \end{aligned} \quad (11)$$

where $\mathbf{C}_{(i,j)} = \text{cov}(\mathbf{y}_{(i)}, \mathbf{y}_{(j)}) = \sum_{u=1}^r \sigma_{u(i,j)} \mathbf{U}_u \mathbf{U}_u^T$ is a covariance matrix between $\mathbf{y}_{(i)}$ and $\mathbf{y}_{(j)}$;

$$\sum_{(1,2,\dots,p-1)} = \text{var} \begin{pmatrix} \mathbf{y}_{(1)} \\ \dots \\ \mathbf{y}_{(p-1)} \end{pmatrix} = \begin{pmatrix} \mathbf{V}_{(1)} & \dots & \mathbf{C}_{(1,p-1)} \\ \dots & \dots & \dots \\ \mathbf{C}_{(p-1,1)} & \dots & \mathbf{V}_{(p-1)} \end{pmatrix}$$

is a $(p-1)n \times (p-1)n$ symmetric matrix.

The new vector in (12) extended from (5) (Zhu, 1995) is independent of all $(p-1)$ conditional variables with the same multiple conditional variance and covariance matrix as in (11).

$$\begin{aligned} & \mathbf{y}_{(p|1,2,\dots,p-1)}^+ \\ &= \mathbf{y}_{(p)} - (\mathbf{C}_{(p,1)} \ \mathbf{C}_{(p,2)} \ \dots \ \mathbf{C}_{(p,p-1)}) \sum_{(1,2,\dots,p-1)}^{-1} \begin{pmatrix} \mathbf{y}_{(1)} - \mathbf{X}\mathbf{b}_{(1)} \\ \dots \\ \mathbf{y}_{(p-1)} - \mathbf{X}\mathbf{b}_{(p-1)} \end{pmatrix}. \end{aligned} \quad (12)$$

Direct use of Eq. (12) to obtain the new independent variable is computationally intensive due to the large size of matrix $\sum_{(1,2,\dots,p-1)}^{-1}$. To reduce computational intensity, another new random vector, which is also independent of $(p-1)$ conditional random variables, can be obtained recursively as follows,

$$\begin{aligned} \mathbf{y}_{(p|1,2,\dots,p-1)}^* &= \mathbf{y}_{(p|1,2,\dots,p-2)}^* - \mathbf{C}_{(p,p-1|1,2,\dots,p-2)} \mathbf{V}_{(p-1|1,2,\dots,p-2)}^{-1} \left(\mathbf{y}_{(p-1|1,2,\dots,p-2)}^* \right. \\ & \quad \left. - \mathbf{X}\mathbf{b}_{(p-1|1,2,\dots,p-2)}^* \right) \\ &= \mathbf{X}\mathbf{b}_{(p|1,2,\dots,p-1)}^* + \sum_{u=1}^r \mathbf{U}_u \mathbf{e}_{u(p|1,2,\dots,p-1)}^*, \end{aligned} \quad (13)$$

where

$$\begin{aligned} & \text{var}(\mathbf{y}_{(p|1,2,\dots,p-1)}^*) \\ &= \mathbf{V}_{(p|1,2,\dots,p-2)} - \mathbf{C}_{(p,p-1|1,2,\dots,p-2)} \mathbf{V}_{(p-1|1,2,\dots,p-2)}^{-1} \mathbf{C}_{(p-1,p|1,2,\dots,p-2)} \\ &= \sum_{u=1}^r \sigma_{u(p|1,2,\dots,p-1)}^{*2} \mathbf{U}_u \mathbf{U}_u^T \end{aligned} \quad (14)$$

$$\begin{aligned} \mathbf{y}_{(p|1,2,\dots,p-2)}^* &= \mathbf{y}_{(p|1,2,\dots,p-3)}^* - \mathbf{C}_{(p,p-2|1,2,\dots,p-3)} \mathbf{V}_{(p-2|1,2,\dots,p-3)}^{-1} \\ & \quad \times \left(\mathbf{y}_{(p-2|1,2,\dots,p-3)}^* - \mathbf{X}\mathbf{b}_{(p-2|1,2,\dots,p-3)}^* \right), \\ \mathbf{y}_{(p-1|1,2,\dots,p-2)}^* &= \mathbf{y}_{(p-1|1,2,\dots,p-3)}^* - \mathbf{C}_{(p-1,p-2|1,2,\dots,p-3)} \mathbf{V}_{(p-2|1,2,\dots,p-3)}^{-1} \\ & \quad \times \left(\mathbf{y}_{(p-2|1,2,\dots,p-3)}^* - \mathbf{X}\mathbf{b}_{(p-2|1,2,\dots,p-3)}^* \right), \end{aligned}$$

$\mathbf{V}_{(p-1|1,2,\dots,p-2)} = \mathbf{V}_{(p-1|1,2,\dots,p-3)} - \mathbf{C}_{(p-1,p-2|1,2,\dots,p-3)} \mathbf{V}_{(p-2|1,2,\dots,p-3)}^{-1} \mathbf{C}_{(p-2,p-1|1,2,\dots,p-3)}$, which is the multivariable conditional variance–covariance matrix for $\mathbf{y}_{(p-1|1,2,\dots,p-2)}$;

$\mathbf{C}_{(p-1,p|1,2,\dots,p-2)} = \mathbf{C}_{(p-1,p|1,2,\dots,p-3)} - \mathbf{C}_{(p-1,p-2|1,2,\dots,p-3)} \mathbf{V}_{(p-2|1,2,\dots,p-3)}^{-1}$
 $\mathbf{C}_{(p-2,p|1,2,\dots,p-3)}$, which is the conditional covariance matrix between $\mathbf{y}_{(p-1|1,2,\dots,p-2)}$
 and $\mathbf{y}_{(p|1,2,\dots,p-2)}$.

It can be shown that the new random vector in formula (13) is independent of $\mathbf{y}_{(1)}$, $\mathbf{y}_{(2)}$, \dots , $\mathbf{y}_{(p-1)}$, with the same variance–covariance matrix as in (11) (see appendix). Thus, $\sigma_{u(p|1,2,\dots,p-1)}^{*2}$ in (14) and $\mathbf{e}_{u(p|1,2,\dots,p-1)}^*$ in (13) can be regarded as equivalent to the conditional variance components $\sigma_{u(p|1,2,\dots,p-1)}^2$ and the conditional effect vector $\mathbf{e}_{u(p|1,2,\dots,p-1)}$.

In practice, variance and covariance components and fixed effects are unknown, so estimation of these parameters is needed for construction of the independent variables. The variance components can be obtained by mixed linear model approaches, such as restricted maximum likelihood (REML) (Patterson and Thompson, 1971) and minimum norm quadratic unbiased estimation (MINQUE) (Rao, 1971; Searle et al., 1992). The covariance components in a mixed linear model can be estimated by the method proposed by Zhu and Weir (1994) and the fixed effects can be obtained by the generalized least-squares estimation method $\hat{\mathbf{b}} = (\mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{X}) \mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{y}$.

2.2. Detecting multivariable conditional variance components and effects

Once the new variables of a target trait independent of multiple component traits have been constructed, mixed linear model approaches can be used to estimate the multivariable conditional variance components (Hartley and Rao, 1967; Patterson and Thompson, 1971; Rao, 1971). Multivariable conditional random effects can be predicted by the best linear unbiased prediction (BLUP) method (Henderson, 1963) if the variance components are known; however, in practice, the variance components usually are estimated from samples, thus, the adjusted unbiased prediction (AUP) (Zhu, 1993) or linear unbiased prediction (LUP) (Zhu and Weir, 1994) are suggested.

3. Application

3.1. Materials and methods

Bolls unit-area⁻¹, lint percentage, and boll size are three important yield components of upland cotton. End-of-season plant mapping data from nine upland cotton commercial cultivars (Shoemaker, 2000) were used to detect multivariable conditional variance components and conditional random effects. The experiments were conducted at the Plant Science Research Center, Mississippi State, MS, in 1997 and 1998. The experimental design was a randomized complete block with six replications in each of 2 years. Plants were grown in a Marietta Silty Clay Loam soil in 1997 and Marietta Sandy Clay Loam soil in 1998. Samples consisting of 25 bolls plot⁻¹ were harvested by hand and ginned to determine lint percentage. The total number of bolls and seed cotton yield within a 3-meter section of row were recorded; boll size was computed as total seed cotton yield/total boll number in the 3 m section; lint yield was calculated based on total seed cotton yield within the 3 m section. A genotype with genotype \times environment interaction model was used for data analysis. The linear model for each trait in terms of matrices and vectors was

Table 1
Estimated unconditional and conditional variance components

	LP ^{a, b}	BN ^b	BS ^b	LY ^b	LY LP	LY BN	LY BS	LY LP &BN	LY LP &BS	LY BN &BS	LY LP& BN&BS
V_G	1.51**	802**	0.29**	398*	208**	351**	211*	213**	201*	110**	34*
V_{GE}	0.66**	495**	0.03**	1144**	646**	249**	1047**	248**	599**	45*	55**
V_e	3.11**	781**	0.03**	2334**	2095**	449**	2222**	198**	2018**	259**	55**
V_P	5.27**	2077**	0.35**	3876**	2949**	1049**	3480**	659**	2818**	414**	143**

*, ** significant at 0.05 and 0.01 probability levels, respectively.

^aLP=lint percentage, BN=boll number per unit area, BS = boll size, LY= lint yield.

^bUnconditional variances. Remaining are conditional variances.

as follows:

$$\mathbf{y} = 1\mu + \mathbf{U}_E\mathbf{e}_E + \mathbf{U}_G\mathbf{e}_G + \mathbf{U}_{GE}\mathbf{e}_{GE} + \mathbf{U}_B\mathbf{e}_B + \mathbf{e} = \mathbf{X}\mathbf{b} + \sum_{u=1}^5 \mathbf{U}_u\mathbf{e}_u, \quad (15)$$

where μ is the fixed population mean; \mathbf{e}_E is the random environmental effect vector, $\mathbf{e}_E \sim N(\mathbf{0}, \sigma_E^2 \mathbf{I}_E)$; \mathbf{U}_E is the design matrix for \mathbf{e}_E ; \mathbf{e}_G is the random genotype effect vector, $\mathbf{e}_G \sim N(\mathbf{0}, \sigma_G^2 \mathbf{I}_G)$; \mathbf{U}_G is the design matrix for \mathbf{e}_G ; \mathbf{e}_{GE} is the random $G \times E$ interaction effect vector, $\mathbf{e}_{GE} \sim N(\mathbf{0}, \sigma_{GE}^2 \mathbf{I}_{GE})$; \mathbf{U}_{GE} is the design matrix for \mathbf{e}_{GE} ; \mathbf{e}_B is the random block effect vector, $\mathbf{e}_B \sim N(\mathbf{0}, \sigma_B^2 \mathbf{I}_B)$; \mathbf{U}_B is the design matrix for \mathbf{e}_B ; and \mathbf{e} is the random error, $\mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$.

Conditional and unconditional variance and covariance components were estimated by MINQUE (1) in which all prior values were set as 1.0 (Zhu, 1989). Conditional and unconditional effects were predicted by AUP (Zhu, 1993). The phenotypic variance was defined as follows: $V_P = V_G + V_{GE} + V_e$ where, $V_G = \sigma_G^2$, $V_{GE} = \sigma_{GE}^2$, and $V_e = \sigma_e^2 \times 1.0 - \sigma_{u(\text{LY}|\text{component}(s))}^2 / \sigma_{u(\text{LY})}^2$ is defined as contribution ratio $CR_{u(\text{component}(s) \rightarrow \text{LY})}$ from single or multiple yield components for u th random effect (Zhu, 1995). $\mathbf{e}_{u(\text{LY})} - \mathbf{e}_{u(\text{LY}|\text{component}(s))}$ is defined as u th net contributed effect $\mathbf{e}_{u(\text{component}(s) \rightarrow \text{LY})}$ from single or joint yield components to lint yield. A resampling (jackknifing) method was applied to calculate the standard error (SE) for each statistics by successive removal of each block within each environment (Miller, 1974). There were six replications in 1997 and 1998 (degrees of freedom = 11). The t -test was used to evaluate the significance of each statistics.

3.2. Results

All correlation coefficients between the constructed random variables of lint yield and conditioned yield component(s) ranged between -0.001 and 0.001 . The results suggested that the constructed random variables were independent of yield component traits, thus they were appropriate for multivariable conditional analysis. Lint yield and all three lint yield components were controlled by both genotypic effects and $G \times E$ interaction effects (Table 1). Lint yield was mainly affected by $G \times E$ interaction effects, while all component traits were mainly affected by genotypic effects. Compared with the unconditional variance components and phenotypic variance, both the conditional variance components and the

Table 2
Contribution ratios (CR) from yield component(s) to lint yield due to different effects

Contribution ratio	LP ^a	BN	BS	LP&BN	LP&BS	BN&BS	LP&BN&BS
$CR_G(\text{component(s)} \rightarrow \text{LY})$	0.48**	0.12	0.47**	0.46**	0.50**	0.73**	0.91**
$CR_{GE}(\text{component(s)} \rightarrow \text{LY})$	0.44**	0.78**	0.09	0.78**	0.48**	0.96**	0.95**
$CR_e(\text{component(s)} \rightarrow \text{LY})$	0.10*	0.81**	0.05	0.92**	0.14**	0.89**	0.98**
$CR_P(\text{component(s)} \rightarrow \text{LY})$	0.24**	0.73**	0.10*	0.83**	0.27**	0.89**	0.96**

*, ** are significant at probability levels of 0.05 and 0.01, respectively.

^aLP=lint percentage, BN=boll number of per unit area, and BS = boll size.

phenotypic variance of lint yield on yield component trait(s) decreased correspondingly. Phenotype contribution ratios of lint percentage and boll size to lint yield were small; however, a large contribution ratio of 47% due to genotypic effect for boll size was detected (Table 2). No significant contribution ratio due to $G \times E$ or residual was detected for boll size. Large contribution ratios due to genotype effect (48%) and $G \times E$ interaction effect (44%), and a small contribution ratio due to residual were detected for lint percentage. Phenotypic contribution ratio of boll number to lint yield was large (73%), while contribution ratio due to $G \times E$ interaction effect and residual were approximately 80%. The contribution ratio for lint percentage and boll size due to each effect was similar to that for lint percentage. The phenotypic contribution ratios for lint percentage and boll number, and for boll size and boll number were greater than 80%. Joint genotypic and $G \times E$ contribution ratios for lint percentage and boll number were similar to those for boll size and boll number. Joint residual contribution ratios for the same two pairs of yield components were approximately 90%. Joint phenotypic contribution ratio and contribution ratio due to each effect from all three component traits were numerically greater than 90%. This suggested that lint percentage, boll number unit-area⁻¹, and boll size can be used to predict lint yield, phenotypically and genetically.

Predicted genotype and $G \times E$ effects of lint yield and contributions of genotype and $G \times E$ effects to lint yield are summarized in Table 3. There was a negative contribution of genotypic effect for FM832 due to boll number but a positive contribution of genotypic effect for FM975. This indicated that the small boll number for FM832 made a negative genotypic contribution to lint yield, large boll number for FM975 made a positive genotype contribution to lint yield, but the remaining cultivars did not show a significant genotypic contribution to lint yield. Genotypic effect of boll size contributed negatively in FM989 but positively in IF1005. Joint genotypic contribution effects due to lint percentage and boll number were negative for DP50 and FM832 but positive for FM975. Joint genotypic contribution effects due to boll size and boll number were negative for FM989 but positive for IF1005. The $G \times E$ contribution effects due to boll number, lint percentage and boll number, boll size and boll number, and three yield components were significant and similar to $G \times E$ effects for FM989 and IF1005 in both years. This indicated that $G \times E$ joint contributed effects to lint yield were mainly from boll number rather than the other two yield components for these two cultivars. Numerically, the joint genotypic and $G \times E$ contribution effects due to the three yield components were similar to the unconditional effects of lint yield for each cultivar.

Table 3

Unconditional genetic effects of lint yield and contribution genetic effects of yield component(s) to lint yield

Cultivar	Main effects (g)							
	G_i	$G_{i(\text{component(s)} \rightarrow \text{LY})}$						
	LY ^a	LP	BN	BS	LP&BN	LP&BS	BN&BS	LP&BN&BS
DP50	−3.49	−22.85	1.07	−2.39	−15.98*	−23.07	5.44	−8.48
FM832	−9.31	−12.75	−25.81**	−11.51	−23.50*	−11.43	−0.34	−3.37
FM963	49.46	27.86	10.61	19.53	24.55	28.82	33.96	43.32
FM975	−0.96	11.87	18.73*	4.77	20.06**	11.52	7.85	8.74
FM989	−11.67	−0.24	−14.37	−8.58†	−8.59	−2.20	−17.94*	−12.52
IF1000	−6.41	5.36	−0.05	−1.59	4.31	4.46	−16.38	−9.88
IF1005	10.75	1.99	17.49	10.05+	9.88	3.07	21.75+	14.35
SG125	−2.80	3.89	−3.34	−1.14	1.80	4.56	−10.87	−5.12
SG501	−25.55	−15.14	−4.33	−9.13	−12.52	−15.73	−23.47	−27.03
$G \times E$ interaction effect (g)								
	GE_i	$GE_{i(\text{component(s)} \rightarrow \text{LY})}$						
	1997							
DP50	−10.93	−45.78**	−9.15	−0.77	−22.56	−43.56**	2.81	−6.96
FM832	12.67	3.19	−4.27	−5.90	−4.05	2.35	10.87	6.86
FM963	42.39+	25.53**	8.33	0.33	13.98	29.51**	30.21	36.76†
FM975	0.75	30.04**	29.65	10.90*	34.33†	26.97**	16.47	18.84
FM989	47.02**	11.20†	45.74**	2.88	45.74**	8.57	37.82*	39.78*
IF1000	29.34†	31.63**	15.44	0.11	20.29	36.63**	23.58	29.53†
IF1005	−49.37*	−32.55**	−43.13*	5.48	−47.35*	−32.81**	−47.18*	−49.49*
SG125	−32.65†	6.45	−22.36	4.39	−17.35	4.21	−39.37	−35.58†
SG501	−16.15	−6.57	2.58	5.63†	−0.06	−9.10	−12.28	−16.69
1998								
DP50	7.54	18.25**	9.94	0.52	6.24	17.04**	2.42	−2.20
FM832	−24.03	−16.02*	−21.84	−11.34*	−21.79	−16.36*	−10.17	−10.44
FM963	17.76	5.31	15.36	2.55	20.61	1.25	11.36	16.24
FM975	−0.88	−14.14	−12.03	−0.21	−13.03	−10.90	−7.31	−7.82
FM989	−58.61*	−11.24	−59.67**	−7.87	−54.25*	−8.76	−56.08*	−52.36*
IF1000	−40.32†	−26.28**	−20.34	1.54	−20.21	−31.22**	−46.44*	−44.74†
IF1005	63.34*	32.14**	64.06**	2.32	60.48*	33.03*	72.98**	67.32**
SG125	25.50†	−2.68	14.78	−6.01	14.88	−0.82	23.94	25.68†
SG501	−13.37	−8.48	−13.07	−4.57	−15.90	−6.02	−13.62	−14.74

†, *, and ** are significant at probability levels of 0.10, 0.05, and 0.01, respectively.

^aLP=lint percentage, BN=boll number of per unit area, BS = boll size, LY= lint yield. Column 1 is the unconditional effect of lint yield. Remaining columns are the contribution effect of yield components on lint yield.

4. Discussion

Many traits, which are determined by multiple component traits, exhibit complex inheritance patterns. Dissecting gene expressions of a complex trait often requires multivariate

analysis. The conditional regression model approach can partition the total variation of a complex trait into variation due to the component traits and residual (Jobson, 1991). Experiments are often conducted across environments with replications and the performance of a complex trait and its component traits could be influenced by environmental conditions, genotypes, and residuals. Conditional regression analysis does not address this complicated situation. The major problems are that the multivariable conditional variance component, or conditional random effects cannot be obtained directly (Zhu, 1995). In this study, Zhu's (1995) method was extended to construct a new vector, which could be used to detect equivalent conditional variance components and conditional random effects. Second, a recursive method was provided to derive the new vector independent of multiple conditional variables for a general mixed linear model to reduce computations based on conditional probability theory. The major advantage of this recursive method is computational speed. There are n^3 total computations required for calculating the inverse of a square matrix of dimension n (Press et al., 1992), while there are p^3n^3 computations required for calculation of the inverse matrix for conditioning on p variables with sample size n by directly using Eq. (12). Therefore, this method can greatly reduce the computational intensity of calculating the inverse of a multi-fold sized matrix when the number of conditioned variables is large. In addition, the program can integrate the Sherman–Morrison–Woodbury method (Press et al., 1992), which can also reduce the calculations required to compute the inverse of a sparse variance–covariance matrix with several variance components (five variance components in the model (15)).

Through multivariate conditional analyses, both phenotypic conditional variance and conditional genotypic effects can be partitioned into several components for a mixed linear model. The results should provide a better understanding of gene expressions of complex traits. In some cases, researchers are interested in the contribution to complex trait from single or multiple component traits. With this in mind, the contribution ratio and the contribution effect defined by Zhu (1995) were also extended. For example, the contribution genetic effects without influence of other related traits is useful for selecting genotypes for use as parents in plant or animal breeding programs.

Numerically, the coefficients of determination (R^2) obtained by both conditional and unconditional multiple linear regression methods should be equivalent, although the mean square errors are different. The phenotypic contribution ratio obtained by this method is also equivalent to the R^2 obtained by multiple regression analyses. R^2 obtained by multiple regression analyses were in agreement with the phenotypic contribution ratio for most cases (Table 4). Thus, if there is only residual variance in a linear model, the multivariate conditional analysis will collapse to a multiple (conditional) linear regression. In addition to multiple conditional analyses, this approach has several other advantages over the traditional linear regression methods: (1) this method can be used to analyze data from more complicated experimental designs, while traditional regression methodology only analyzes phenotypic data; (2) this method can partition the phenotypic contribution ratio into different contribution ratios for a specific mixed linear model, while the traditional regression methods cannot; and (3) the contribution effects, which are useful in selection studies, can be predicted.

The models in this application section can be extended to other mixed linear models. Based on Cockerham's (1980) genetic model, for example, the model can be extended to

Table 4

Coefficients of determination (R^2) obtained by conditional and unconditional linear regression analyses

	LP ^a	BN	BS	LP&BN	LP&BS	BN&BS	LP&BN&BS
Unconditional ^b	0.24	0.68	0.02	0.72	0.24	0.91	0.97
Conditional ^c	0.24	0.68	0.02	0.72	0.24	0.90	0.96

^aLP=lint percentage, BN=boll number per unit area, and BS = boll size.^bUsing regular linear regression model.^cUsing Jobson (1991) model.

additive-dominance, additive dominance additive \times additive, or to animal genetic models (Zhu and Weir, 1994).

The detection of conditional variances and conditional effects is based on the constructed random variable, which is phenotypically independent of the component trait(s); however, it still remains unproven if each effect and variance component is theoretically independent. Based on the analysis of covariance components by the approach proposed by Zhu (1992), the correlation coefficient for each effect between the new constructed variables and each of the yield component traits was approximately zero and insignificant (data not presented). In addition, this method for constructing a new independent random variable for multivariable conditional analysis is based on the assumption of normal distribution. This method may also be applicable to other types of distributions. Based on the central limit theorem, other distributions can be converted to normal distributions if large sample sizes are used. In some cases, estimating multivariable conditional fixed effects is also important. One alternative way to do this is to treat the fixed effects (i.e., treatment effects) as random and conditional fixed effects could be predicted by the LUP (Zhu and Weir, 1994) or AUP (Zhu, 1993) approach.

Appendix

Let $\mathbf{y}_{(1)}, \mathbf{y}_{(2)}, \dots, \mathbf{y}_{(p)}$ be jointly normally distributed for $p \geq 2$.

$$\begin{aligned} \mathbf{y}_{(p|1,2,\dots,p-1)}^* &= \mathbf{y}_{(p|1,2,\dots,p-2)}^* - \mathbf{C}_{(p,p-1|1,2,\dots,p-2)} \mathbf{V}_{(p-1|1,2,\dots,p-2)}^{-1} \\ &\quad \times \left(\mathbf{y}_{(p-1|1,2,\dots,p-2)}^* - \mathbf{X} \mathbf{b}_{(p-1|1,2,\dots,p-2)}^* \right). \end{aligned}$$

Then, $\text{var}(\mathbf{y}_{(p|1,2,\dots,p-1)}^*) = \text{var}(\mathbf{y}_{(p|1,2,\dots,p-1)})$ and $\text{cov}(\mathbf{y}_{(p|1,2,\dots,p-1)}^*, \mathbf{y}_{(k)}) = \mathbf{0}$ for $k = 1, 2, \dots, (p-1)$ are true.

Prove by induction:

(1) $p = 2$.

Let $f(\cdot)$ $f(\cdot, \cdot)$ be probability density function (PDF) and joint PDF, respectively.

Since $f(\mathbf{y}_{(2|1)}) = f(\mathbf{y}_1, \mathbf{y}_2)/f(\mathbf{y}_1)$, then $\mathbf{V}_{(2|1)} = \text{var}(\mathbf{y}_{(2|1)}) = \mathbf{V}_{(2)} - \mathbf{C}_{(2,1)}\mathbf{V}_{(1)}^{-1}\mathbf{C}_{(1,2)}$.

$$\begin{aligned}\mathbf{V}_{(2|1)}^* &= \text{var}(\mathbf{y}_{(2|1)}^*) = \text{var}(\mathbf{y}_{(2)} - \mathbf{C}_{(2,1)}\mathbf{V}_{(1)}^{-1}(\mathbf{y}_{(1)} - \mathbf{X}\mathbf{b}_{(1)})) \\ &= \text{var}(\mathbf{y}_{(2)}) + \mathbf{C}_{(2,1)}\mathbf{V}_{(1)}^{-1}\text{var}(\mathbf{y}_{(1)} - \mathbf{X}\mathbf{b}_{(1)})\mathbf{V}_{(1)}^{-1}\mathbf{C}_{(1,2)} \\ &\quad - 2\text{cov}(\mathbf{y}_{(2)}, \mathbf{C}_{(2,1)}\mathbf{V}_{(1)}^{-1}(\mathbf{y}_{(1)} - \mathbf{X}\mathbf{b}_{(1)})) \\ &= \mathbf{V}_{(2)} - \mathbf{C}_{(2,1)}\mathbf{V}_{(1)}^{-1}\mathbf{C}_{(1,2)},\end{aligned}$$

so $\mathbf{V}_{(2|1)}^* = \mathbf{V}_{(2|1)}$,

$\text{cov}(\mathbf{y}_{2|1}^*, \mathbf{y}_1) = \text{cov}\{\mathbf{y}_2 - \mathbf{C}_{21}\mathbf{V}_1^{-1}(\mathbf{y}_1 - \mathbf{X}\mathbf{b}_1), \mathbf{y}_1\} = \mathbf{0}$, which means new vector $\mathbf{y}_{2|1}^*$ is independent of \mathbf{y}_1 with the same conditional variance–covariance matrix of $\mathbf{V}_{(2|1)}$.

It is true for $p = 2$.

(2) $p = 3$.

The PDF for conditional random vector $\mathbf{y}_{(3)}$ on $\mathbf{y}_{(1)}$ and $\mathbf{y}_{(2)}$ can be expressed as follows:

$$f(\mathbf{y}_{(3|1,2)}) = f(\mathbf{y}_{(2|1)}, \mathbf{y}_{(3|1)})/f(\mathbf{y}_{(2|1)}).$$

Following (1), we can obtain $\mathbf{V}_{(3|1,2)} = \text{var}(\mathbf{y}_{(3|1,2)}) = \mathbf{V}_{(3|1)} - \mathbf{C}_{(3,2|1)}\mathbf{V}_{(2|1)}^{-1}\mathbf{C}_{(2,3|1)}$.

Suppose two new vectors $\mathbf{y}_{(2|1)}^*$ and $\mathbf{y}_{(3|1)}^*$ are constructed from (1). From results of (1), we get

$$\begin{aligned}\mathbf{V}_{(3|1,2)}^* &= \text{var}(\mathbf{y}_{(3|1,2)}^*) = \text{var}\left(\mathbf{y}_{(3|1)}^* - \mathbf{C}_{(3,2|1)}\mathbf{V}_{(2|1)}^{-1}\left(\mathbf{y}_{(2|1)}^* - \mathbf{X}\mathbf{b}_{(2|1)}^*\right)\right) \\ &= \mathbf{V}_{(3|1)}^* - \mathbf{C}_{(3,2|1)}^*\mathbf{V}_{(2|1)}^{*-1}\mathbf{C}_{(2,3|1)}^* \\ &= \mathbf{V}_{(3|1)} - \mathbf{C}_{(3,2|1)}^*\mathbf{V}_{(2|1)}^{-1}\mathbf{C}_{(2,3|1)}^*,\end{aligned}$$

since

$$\begin{aligned}\mathbf{C}_{(2,3|1)}^* &= \text{cov}(\mathbf{y}_{(2|1)}^*, \mathbf{y}_{(3|1)}^*) \\ &= \text{cov}\{\mathbf{y}_{(2)} - \mathbf{C}_{(2,1)}\mathbf{V}_{(1)}^{-1}(\mathbf{y}_{(1)} - \mathbf{X}\mathbf{b}_{(1)}), \mathbf{y}_{(3)} - \mathbf{C}_{(3,1)}\mathbf{V}_{(1)}^{-1}(\mathbf{y}_{(1)} - \mathbf{X}\mathbf{b}_{(1)})\} \\ &= \mathbf{C}_{(2,3)} - \mathbf{C}_{(2,1)}\mathbf{V}_{(1)}^{-1}\mathbf{C}_{(1,3)},\end{aligned}$$

so $\mathbf{C}_{(2,3|1)}^* = \text{cov}(\mathbf{y}_{2|1}, \mathbf{y}_{3|1}) = \mathbf{C}_{(2,3|1)}$, which means that covariance matrix of new vectors of $\mathbf{y}_{(2|1)}^*, \mathbf{y}_{(3|1)}^*$ is equivalent to conditional covariance matrix between $\mathbf{y}_{(2|1)}$ and $\mathbf{y}_{(3|1)}$.

$$\mathbf{C}_{(2,3|1)}^T = \mathbf{C}_{(3,2|1)}.$$

Thus,

$$\mathbf{V}_{(3|1,2)}^* = \mathbf{V}_{(3|1)} - \mathbf{C}_{(3,2|1)}\mathbf{V}_{(2|1)}^{-1}\mathbf{C}_{(2,3|1)} = \mathbf{V}_{(3|1,2)},$$

$$\begin{aligned}\text{cov}(\mathbf{y}_{(3|1,2)}^*, \mathbf{y}_{(1)}) &= \text{cov}\left\{\left(\mathbf{y}_{(3|1)}^* - \mathbf{C}_{(3,2|1)}\mathbf{V}_{(2|1)}^{-1}(\mathbf{y}_{(2|1)}^* - \mathbf{X}\mathbf{b}_{(2|1)}^*)\right), \mathbf{y}_{(1)}\right\} \\ &= \text{cov}(\mathbf{y}_{(3|1)}^*, \mathbf{y}_{(1)}) - \mathbf{C}_{(3,2|1)}\mathbf{V}_{(2|1)}^{-1}\text{cov}(\mathbf{y}_{(2|1)}^*, \mathbf{y}_{(1)}) = \mathbf{0}.\end{aligned}$$

Similarly, it can be shown that $\text{cov}(\mathbf{y}_{(3|1,2)}^*, \mathbf{y}_{(2)}) = \mathbf{0}$.

Thus, vector $\mathbf{y}_{(3|1,2)}^*$ is independent of vectors $\mathbf{y}_{(1)}$ and $\mathbf{y}_{(2)}$ with the same conditional variance–covariance matrix as for $\mathbf{V}_{(3|1,2)}$.

It is true for $p = 3$.

(3) Suppose for $2 \leq t \leq p-1$ ($p \geq 4$), the result is established such that $\text{var}(\mathbf{y}_{(p-1|1,2,\dots,p-2)}^*) = \text{var}(\mathbf{y}_{(p-1|1,2,\dots,p-2)})$, $\text{cov}(\mathbf{y}_{(p-1|1,2,\dots,p-2)}^*, \mathbf{y}_{(k)}) = \mathbf{0}$ for $k = 1, 2, \dots, (p-2)$, and $\text{cov}(\mathbf{y}_{(p-1|1,2,\dots,p-3)}^*, \mathbf{y}_{(p-2|1,2,\dots,p-3)}^*) = \mathbf{C}_{(p-1,p-2|1,2,\dots,p-3)}$ are true, we will derive the case of p random vectors.

The PDF of conditional random vector $\mathbf{y}_{(p|1,2,\dots,p-1)}$ can be expressed as follows:

$$f(\mathbf{y}_{(p|1,2,\dots,p-1)}) = f(\mathbf{y}_{(p|1,2,\dots,p-2)}, \mathbf{y}_{(p-1|1,2,\dots,p-2)}) / f(\mathbf{y}_{(p-1|1,2,\dots,p-2)}), \quad \text{then} \\ \text{var}(\mathbf{y}_{(p|1,2,\dots,p-1)}) = \mathbf{V}_{(p|1,2,\dots,p-1)} = \mathbf{V}_{(p|1,2,\dots,p-2)} - \mathbf{C}_{(p,p-1|1,2,\dots,p-2)} \mathbf{V}_{(p-1|1,2,\dots,p-2)}^{-1} \mathbf{C}_{(p-1,p|1,2,\dots,p-2)}.$$

Let $\mathbf{y}_{(p|1,2,\dots,p-2)}^*$ and $\mathbf{y}_{(p-1|1,2,\dots,p-2)}^*$ be available, thus, both $\mathbf{y}_{(p|1,2,\dots,p-2)}^*$ and $\mathbf{y}_{(p-1|1,2,\dots,p-2)}^*$ are independent of $\mathbf{y}_{(k)}$ for $k = 1, 2, \dots, (p-2)$.

$$\mathbf{y}_{(p|1,2,\dots,p-1)}^* = \mathbf{y}_{(p|1,2,\dots,p-2)}^* - \mathbf{C}_{(p,p-1|1,2,\dots,p-2)} \mathbf{V}_{(p-1|1,2,\dots,p-2)}^{-1} \left(\mathbf{y}_{(p-1|1,2,\dots,p-2)}^* - \mathbf{Xb}_{(p-1|1,2,\dots,p-2)}^* \right),$$

$$\begin{aligned} \mathbf{V}_{(p|1,2,\dots,p-1)}^* &= \text{var}(\mathbf{y}_{(p|1,2,\dots,p-1)}^*) \\ &= \text{var} \left(\mathbf{y}_{(p|1,2,\dots,p-2)}^* - \mathbf{C}_{(p,p-1|1,2,\dots,p-2)} \mathbf{V}_{(p-1|1,2,\dots,p-2)}^{-1} \left(\mathbf{y}_{(p-1|1,2,\dots,p-2)}^* - \mathbf{Xb}_{(p-1|1,2,\dots,p-2)}^* \right) \right) \\ &= \mathbf{V}_{(p|1,2,\dots,p-2)} - \mathbf{C}_{(p,p-1|1,2,\dots,p-2)}^* \mathbf{V}_{(p-1|1,2,\dots,p-2)}^{-1} \mathbf{C}_{(p-1,p|1,2,\dots,p-2)}^*. \end{aligned}$$

Note that

$$\begin{aligned} \mathbf{C}_{(p,p-1|1,2,\dots,p-2)}^* &= \text{cov} \left(\mathbf{y}_{(p|1,2,\dots,p-2)}^*, \mathbf{y}_{(p-1|1,2,\dots,p-2)}^* \right) \\ &= \text{cov} \left(\mathbf{y}_{(p|1,2,\dots,p-3)}^* - \mathbf{C}_{(p,p-2|1,2,\dots,p-3)} \mathbf{V}_{(p-2|1,2,\dots,p-3)}^{-1} \left(\mathbf{y}_{(p-2|1,2,\dots,p-3)}^* - \mathbf{Xb}_{(p-2|1,2,\dots,p-3)}^* \right), \right. \\ &\quad \left. \mathbf{y}_{(p-1|1,2,\dots,p-3)}^* - \mathbf{C}_{(p-1,p-2|1,2,\dots,p-3)} \mathbf{V}_{(p-2|1,2,\dots,p-3)}^{-1} \left(\mathbf{y}_{(p-2|1,2,\dots,p-3)}^* - \mathbf{Xb}_{(p-2|1,2,\dots,p-3)}^* \right) \right) \\ &= \text{cov} \left(\mathbf{y}_{(p|1,2,\dots,p-3)}^*, \mathbf{y}_{(p-1|1,2,\dots,p-3)}^* \right) \\ &\quad - \text{cov} \left(\mathbf{C}_{(p,p-2|1,2,\dots,p-3)} \mathbf{V}_{(p-2|1,2,\dots,p-3)}^{-1} \mathbf{y}_{(p-2|1,2,\dots,p-3)}^*, \mathbf{y}_{(p-1|1,2,\dots,p-3)}^* \right) \\ &\quad - \text{cov} \left(\mathbf{y}_{(p|1,2,\dots,p-3)}^*, \mathbf{C}_{(p-1,p-2|1,2,\dots,p-3)} \mathbf{V}_{(p-2|1,2,\dots,p-3)}^{-1} \mathbf{y}_{(p-2|1,2,\dots,p-3)}^* \right) \\ &\quad + \text{cov} \left(\mathbf{C}_{(p,p-2|1,2,\dots,p-3)} \mathbf{V}_{(p-2|1,2,\dots,p-3)}^{-1} \mathbf{y}_{(p-2|1,2,\dots,p-3)}^*, \right. \\ &\quad \left. \mathbf{C}_{(p-1,p-2|1,2,\dots,p-3)} \mathbf{V}_{(p-2|1,2,\dots,p-3)}^{-1} \mathbf{y}_{(p-2|1,2,\dots,p-3)}^* \right) \end{aligned}$$

$$\begin{aligned}
&= \mathbf{C}_{(p,p-1|1,2,\dots,p-3)} - \mathbf{C}_{(p,p-2|1,2,\dots,p-3)} \mathbf{V}_{(p-2|1,2,\dots,p-3)}^{-1} \\
&\quad \times \text{cov} \left(\mathbf{y}_{(p-2|1,2,\dots,p-3)}^*, \mathbf{y}_{(p-1|1,2,\dots,p-3)}^* \right) \\
&\quad - \text{cov} \left(\mathbf{y}_{(p|1,2,\dots,p-3)}^*, \mathbf{y}_{(p-2|1,2,\dots,p-3)}^* \right) \mathbf{V}_{(p-2|1,2,\dots,p-3)}^{-1} \mathbf{C}_{(p-2,p-1|1,2,\dots,p-3)} \\
&\quad + \mathbf{C}_{(p,p-2|1,2,\dots,p-3)} \mathbf{V}_{(p-2|1,2,\dots,p-3)}^{-1} \text{var}(\mathbf{y}_{(p-2|1,2,\dots,p-3)}^*) \\
&\quad \times \mathbf{V}_{(p-2|1,2,\dots,p-3)}^{-1} \mathbf{C}_{(p-2,p-1|1,2,\dots,p-3)} \\
&= \mathbf{C}_{(p,p-1|1,2,\dots,p-3)} - \mathbf{C}_{(p,p-2|1,2,\dots,p-3)} \mathbf{V}_{(p-2|1,2,\dots,p-3)}^{-1} \mathbf{C}_{(p-2,p-1|1,2,\dots,p-3)} \\
&\quad - \mathbf{C}_{(p,p-2|1,2,\dots,p-3)} \mathbf{V}_{(p-2|1,2,\dots,p-3)}^{-1} \mathbf{C}_{(p-2,p-1|1,2,\dots,p-3)} \\
&\quad + \mathbf{C}_{(p,p-2|1,2,\dots,p-3)} \mathbf{V}_{(p-2|1,2,\dots,p-3)}^{-1} \\
&\quad \times \mathbf{V}_{(p-2|1,2,\dots,p-3)} \mathbf{V}_{(p-2|1,2,\dots,p-3)}^{-1} \mathbf{C}_{(p-2,p-1|1,2,\dots,p-3)} \\
&= \mathbf{C}_{(p,p-1|1,2,\dots,p-3)} - \mathbf{C}_{(p,p-2|1,2,\dots,p-3)} \mathbf{V}_{(p-2|1,2,\dots,p-3)}^{-1} \mathbf{C}_{(p-2,p-1|1,2,\dots,p-3)} \\
&\quad - \mathbf{C}_{(p,p-2|1,2,\dots,p-3)} \mathbf{V}_{(p-2|1,2,\dots,p-3)}^{-1} \mathbf{C}_{(p-2,p-1|1,2,\dots,p-3)} \\
&\quad + \mathbf{C}_{(p,p-2|1,2,\dots,p-3)} \mathbf{V}_{(p-2|1,2,\dots,p-3)}^{-1} \mathbf{C}_{(p-2,p|1,2,\dots,p-3)} \\
&= \mathbf{C}_{(p,p-1|1,2,\dots,p-3)} - \mathbf{C}_{(p,p-2|1,2,\dots,p-3)} \mathbf{V}_{(p-2|1,2,\dots,p-3)}^{-1} \mathbf{C}_{(p-2,p-1|1,2,\dots,p-3)} \\
&= \text{cov}(\mathbf{y}_{(p|1,2,\dots,p-2)}, \mathbf{y}_{(p-1|1,2,\dots,p-2)}) \\
&= \mathbf{C}_{(p,p-1|1,2,\dots,p-2)}
\end{aligned}$$

and

$$\mathbf{C}_{(p,p-1|1,2,\dots,p-3)}^T = \mathbf{C}_{(p-1,p|1,2,\dots,p-2)}.$$

Thus,

$$\begin{aligned}
&\mathbf{V}_{(p|1,2,\dots,p-1)}^* \\
&= \mathbf{V}_{(p|1,2,\dots,p-2)} - \mathbf{C}_{(p,p-1|1,2,\dots,p-2)}^* \mathbf{V}_{(p-1|1,2,\dots,p-2)}^{-1} \mathbf{C}_{(p-1,p|1,2,\dots,p-2)}^* \\
&= \mathbf{V}_{(p|1,2,\dots,p-2)} - \mathbf{C}_{(p,p-1|1,2,\dots,p-2)} \mathbf{V}_{(p-1|1,2,\dots,p-2)}^{-1} \mathbf{C}_{(p-1,p|1,2,\dots,p-2)} \\
&= \text{var}(\mathbf{y}_{(p|1,2,\dots,p-1)}) = \mathbf{V}_{(p|1,2,\dots,p-1)}, \\
&\text{cov}(\mathbf{y}_{(p|1,2,\dots,p-1)}^*, \mathbf{y}_{(k)}) \\
&= \text{cov} \left(\mathbf{y}_{(p|1,2,\dots,p-2)}^* - \mathbf{C}_{(p,p-1|1,2,\dots,p-2)} \mathbf{V}_{(p-1|1,2,\dots,p-2)}^{-1} \left(\mathbf{y}_{(p-1|1,2,\dots,p-2)}^* \right. \right. \\
&\quad \left. \left. - \mathbf{X} \mathbf{b}_{(p-1|1,2,\dots,p-2)}^* \right), \mathbf{y}_{(k)} \right) \\
&= \text{cov} \left(\mathbf{y}_{(p|1,2,\dots,p-2)}^* - \mathbf{C}_{(p,p-1|1,2,\dots,p-2)} \mathbf{V}_{(p-1|1,2,\dots,p-2)}^{-1} \right. \\
&\quad \left. \times \mathbf{y}_{(p-1|1,2,\dots,p-2)}^*, \mathbf{y}_{(k)} \right)
\end{aligned}$$

$$\begin{aligned}
&= \text{cov} \left(\mathbf{y}_{(p|1,2,\dots,p-2)}^*, \mathbf{y}_{(k)} \right) - \mathbf{C}_{(p,p-1|1,2,\dots,p-2)} \mathbf{V}_{(p-1|1,2,\dots,p-2)}^{-1} \\
&\quad \times \text{cov} \left(\mathbf{y}_{(p-1|1,2,\dots,p-2)}^*, \mathbf{y}_{(k)} \right) \\
&= \mathbf{0} - \mathbf{C}_{(p,p-1|1,2,\dots,p-2)} \mathbf{V}_{(p-1|1,2,\dots,p-2)}^{-1} \mathbf{0} \\
&= \mathbf{0}.
\end{aligned}$$

Since $\mathbf{y}_{(p|1,2,\dots,p-1)}^*$ can also be expressed differently such that

$\mathbf{y}_{(p|1,2,\dots,p-1)}^* = \mathbf{y}_{(p|2,\dots,p-1)}^* - \mathbf{C}_{(p,1|2,\dots,p-1)} \mathbf{V}_{(1|2,\dots,p-1)}^{-1} \left(\mathbf{y}_{(1|2,\dots,p-1)}^* - \mathbf{Xb}_{(1|2,\dots,p-1)}^* \right)$, which has the same conditional variance–covariance matrix as of $\mathbf{V}_{(p|1,2,\dots,p-1)}$ and is independent of vector $\mathbf{y}_{(p-1)}$.

Therefore, it turns out that it is true for the case of p ($p > 3$) random vectors.

References

- Atchley, W.R., Zhu, J., 1997. Developmental quantitative genetics, conditional epigenetic variability and growth in mice. *Genetics* 147, 765–776.
- Ball, R.A., McNew, R.W., Vories, E.D., Keisling, T.C., Purcell, L.C., 2001. Path analyses of population density effects on short-season soybean yield. *Agro. J.* 93, 187–195.
- Bora, G.C., Gupta, S.N., Tomer, Y.S., Singh, S., 1998. Genetic variability, correlation and path analysis in faba bean (*Vicia faba*). *Indian J. Agric. Sci.* 68, 212–214.
- Cockerham, C.C., 1980. Random and fixed effect in plant genetics. *Theor. Appl. Genet.* 56, 119–131.
- Cramer, C.S., Wehner, T.C., 2000. Path analysis of the correlation between fruit number and plant traits of cucumber populations. *Hort. Sci.* 35, 708–711.
- Dilday, R.H., Mgonja, M.A., Amonsilpa, S.A., Collins, F.C., Wells, B.R., 1990. Plant height vs. mesocotyl and coleoptile elongation in rice: linkage or pleiotropism. *Crop Sci.* 30, 815–818.
- Graybill, F.A., 1976. *Theory and Application of Linear Model*, Duxbury Press, Boston, MA.
- Hartley, H.O., Rao, J.N.K., 1967. Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika* 54, 93–108.
- Henderson, C.R., 1963. Selection index and expected genetic advance. Hanson, W.D., Robinson, H.F. (Eds.), *Statistical Genetics and Plant Breeding*, vol. 982. National Academic Science and National Research Council Publishers, pp. 141–163.
- Jobson, J.D., 1991. *Applied Multivariate Data Analysis*, vol II: categorical and multivariate methods, Springer, Berlin.
- Kebede, H., Subudhi, P.K., Rosenow, D.T., Nguyen, H.T., 2001. Quantitative trait loci influencing drought tolerance in grain sorghum (*Sorghum bicolor* L. Moench). *Theor. Appl. Genet.* 103, 266–276.
- Krzanowski, W.J., 1988. *Principles of Multivariate Analysis*, Oxford Science Publications, Oxford.
- Miller, R.G., 1974. The jackknife: a review. *Biometrika* 61, 1–15.
- Patterson, H.D., Thompson, R., 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika* 58, 545–554.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. *Numerical Recipes in C: the Art of Scientific Computing*, second ed. Cambridge University Press, Cambridge.
- Rao, C.R., 1971. Estimation of variance and covariance components MINQUE theory. *J. Mutivariate Anal.* 1, 257–275.
- Samonte, S.O.P., Wilson, L.T., McClung, A.M., 1998. Path analyses of yield and yield-related traits of fifteen diverse rice genotypes. *Crop Sci.* 38, 1130–1136.
- Searle, S.R., Casella, G., McCulloch, C.E., 1992. *Variance Components*, Wiley, New York, NY.
- Shoemaker, D.B., 2000. Genetic analyses of agronomic traits of selected American and Australian cotton genotypes and their F₂ hybrids. Ph.D. Dissertation, Mississippi State University, Mississippi State, MS.

- Wright, S., 1920. The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs. *Proc. Natl. Acad. Sci. USA* 6, 320–332.
- Wu, J., Zhu, J., Xu, F., Ji, D., 1995. Analysis of genetic effect \times environment interactions for yield traits in upland cotton. *Heredity* 17 (5), 1–4 (in Chinese).
- Yan, J., Zhu, J., He, C., Benmoussa, M., Wu, P., 1998. Quantitative trait loci analysis for developmental behavior of tiller number in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 97, 267–274.
- Zhu, J., 1989. Estimation of genetic variance components in the general mixed model. Ph.D. Dissertation, North Carolina State University, Raleigh, NC.
- Zhu, J., 1993. Methods of predicting genotype value and heterosis for offspring of hybrids. *J. Biomath.* 8 (1), 32–44 (in Chinese).
- Zhu, J., 1995. Analysis of conditional effects and variance components in developmental genetics. *Genetics* 141, 1633–1699.
- Zhu, J., Weir, B.S., 1994. Analysis of cytoplasmic and maternal effects: I. a genetic model for diploid plant seeds and animals. *Theor. Appl. Genet.* 89, 153–159.